



Aus Dieter E. Zimmer:
Die Elektrifizierung der Sprache
Zürich: Haffmans, 1990, S.203-237
Als Taschenbuch München: Heyne, 1997
S.211-248
© 1990 Haffmans Verlag
und Dieter E. Zimmer

Mr. Searle im Chinesischen Zimmer

Über Computer, Gehirne und Geist

EINE PROVOKATION feiert Geburtstag. Vor nunmehr zehn Jahren löste sie eine kontroverse Debatte aus, die anders als die meisten wissenschaftlichen Debatten nach dem Austausch der Argumente und Daten nicht im Nu erledigt war. Sie zieht sich bis heute hin und macht keinerlei Anstalten, sich zu legen.

Vordergründig geht es um Chinesische Zimmer und Chinesische Turnhallen, um Schnellrestaurants, um sprechende Maschinen, um Computer und wieviel Intelligenz sie eines Tages ihr eigen nennen könnten – und in Wahrheit bei alledem um die Letzten Dinge, jene, die Leidenschaften wecken: Was ist der menschliche Geist? Kann es eines Tages eine Maschine geben, die Geist hat?

Was die Kontroverse in Gang setzte, war eine Herausforderung an die junge Disziplin der Künstlichen Intelligenz. Das menschliche Geistorgan, so lautete sie, funktioniert nicht wie ein Computer, und folglich könne ein Computer es auch nie und nimmer duplizieren.

Der Provokateur war John Searle, seit 1959 Philosophieprofessor an der Universität Kalifornien in Berkeley und heute 57 Jahre alt. Seine Provokation bestand in einem Aufsatz, den er — nach jahrelanger Erprobung seiner Argumente — in der hervorragenden wissenschaftlichen Zeitschrift *The Behavioral and Brain Sciences* erscheinen ließ. Dieses Journal sieht sein Ziel darin, prominente wissenschaftliche Theorien sachverständiger Kritik auszusetzen; es druckt Aufsätze nicht pur, sondern zusammen mit Dutzenden von ausführlichen Kommentaren aus der Fachszene, die nicht selten eigene Forschungsbeiträge sind. Schon damals war auszumachen, daß Searles Aufsatz eingeschlagen hatte. Seitdem hat es zustimmende

Elaborationen und Widerlegungsversuche zuhauf gegeben; alles zusammen würde Bände füllen. Im März 1990 brach die Debatte in der Zeitschrift *Spektrum der Wissenschaft* noch einmal auf, und alles war wie am ersten Tag — keine Seite hatte im mindesten zurückgesteckt.

Die Künstliche Intelligenz (international meist AI genannt, Artifizielle Intelligenz) ist jener Teil der Informatik, der — nicht ohne praktische Erfolge — den Computer dazu zu bringen sucht, gewisse menschliche Leistungen nachzuahmen, die wir, wenn ein Mensch aus Fleisch und Blut sie vollbrächte, «intelligent» zu nennen uns nicht genieren würden.

Wie Searle es sieht, kommt der Anspruch der AI in einer schwachen und einer starken Form. Die «schwache AI» behauptet nur, daß der Computer ein brauchbares Instrument zur Erforschung menschlichen Denkvermögens darstelle. Ob er dazu taugt, ist längst keine Frage mehr. Vor allem zwingt er, vollkommen explizit zu machen, welche Teilleistungen in eine sonst eher global betrachtete Intelligenzleistung eingehen. («Lesen» ist schön und gut; aber was muß im einzelnen geschehen, und in welcher Reihenfolge muß es geschehen, bis die Schriftzeichen auf dem Papier zu einer Bedeutung im Kopf geworden sind?) Auch Searle hatte und hat keinen Zweifel daran, daß sich manche geistige Tat auf dem Computer simulieren läßt; und daß man auf dem Weg zu einer perfekten Simulation viel darüber lernen könne, wie der menschliche Geist funktionieren müsse. Er warnt nur davor, solche Simulationen ohne weiteres schon für das Eigentliche zu halten, *the real thing*: So wenig, wie ein simuliertes Unwetter naß mache oder die Simulation eines Verdauungssystems irgend etwas verdaue, so wenig sei ein simulierter Denkprozeß dasselbe wie menschliches Denken.

Damit aber legte er sich mit dem weitergehenden Anspruch mancher AI-Forscher an, der ihre Disziplin auf Trab hält: daß der Computer, wenn man ihn nur mit den richtigen Programmen füttere, Geist nicht nur simuliere, sondern ihn ohne Wenn und Aber besitze. Die Implikation dieser «starken AI»: «Geist», «Verstand», «Denkvermögen», «bewußte Intelligenz» und so weiter sei nichts anderes als ein Computerprogramm, das wie alle Computerprogramme an keine bestimmte Hardware gebunden ist und folglich auf Digitalrechnern, auf Computern aus alten Bierdosen, auf den unvorstellbaren Denkkapparaten extraterrestrischer Wesen laufen kann, wie es auf biologischen Gehirnen läuft. In den bewegten Worten von Douglas R. Hofstadter: «Der Geist existiert in Gehirnen und eines Tages vielleicht auch in programmierten Maschinen. Wenn es zu solchen Maschinen kommen sollte, dann werden sich ihre kausalen Kräfte nicht aus dem Material ergeben, aus dem sie bestehen, sondern aus ihrem Design und aus den Programmen, die in ihnen laufen. Und daß sie jene kausalen Kräfte besitzen, werden wir daher wissen, daß wir mit ihnen sprechen und uns sorgfältig anhören, was sie zu sagen haben.»

Ob man einer Maschine zugestehen muß, sie besitze menschliche Intelligenz, dafür hat der Computertheoretiker Alan Turing einst einen be-

rühmten Test ersonnen. Ein Mensch sitzt in einem Zimmer, im Nebenzimmer steht ein Computer. Der Mensch weiß aber nicht, ob es ein Computer ist oder ein Mensch. Um es zu entscheiden, stellt er dem Nebenzimmer — über eine Schreibkonsole — Fragen über Fragen. Und wenn der Mensch an den Antworten nicht zu erkennen vermag, daß deren Urheber eine Maschine war, dann hat sie den Turing-Test bestanden. Sie besitze menschliche Intelligenz!

Natürlich gab es zu Turings Zeiten keine Maschine, der man alle Fragen stellen konnte, die sich einem Menschen stellen lassen, und auch heute ist sie nicht in Sicht. Dennoch ist dieser Test gar nicht so utopisch. Einige spezielle Denkleistungen, die einmal dem Menschen vorbehalten schienen, hat der Computer inzwischen tatsächlich gemeistert. Wer heute beim Schachspiel im unklaren gelassen würde, ob ein Mensch oder ein Computer die Gegenzüge ersinnt, hätte größte Mühe, den Computer zu identifizieren. Aber es ist ein rein behavioristischer Test. Der Behaviorismus war jene einstmals mächtige psychologische Schule, die sich programmatisch nur für das interessierte, was in die Seele an «Reizen» eingeht und an «Verhalten» herauskommt, aber nicht für das, was in ihr vorgeht. Das Innere schien dem Behaviorismus ein für allemal unzugänglich und sollte darum ganz aus dem Spiel bleiben: eine Black Box. Innere Dispositionen ließen sich als Dispositionen zu bestimmtem Verhalten beschreiben, und schon wäre man die unbequeme, objektiver Beobachtung verschlossene Domäne des subjektiven Erlebens los. Der Mensch sei sein Verhalten, basta.

Genau das ist auch die Prämisse des Turing-Tests. Ob sich im Computer, wenn er Verstand simuliert, etwas ähnliches abspielt wie im Menschen, interessiert ihn nicht. Nur auf den Output kommt es ihm an. Wenn der Output eines Computers sich nicht von dem eines Menschen unterscheiden läßt, dann spricht er dem Computer menschlichen Verstand zu. Oder wie der Philosoph Paul Churchland es formulierte: «Wenn Maschinen es schaffen, alle unsere inneren kognitiven Tätigkeiten bis ins letzte Detail zu simulieren, dann wäre es geradezu ein neuer Rassismus, ihnen den Status echter Personen vorzuenthalten.»

Searle nun erdachte ein Gedankenexperiment, das listig an den Turing-Test anknüpft. Es gibt AI-Programme, die menschliches Sprachverstehen simulieren. Am bekanntesten ist SAM von Roger Schank, einem AI-Forscher der Universität Princeton. Man erzählt SAM (vorerst schriftlich) kurze Geschichten, und er beantwortet Fragen über sie. Besonderen Wert legte Schank darauf, daß SAM nicht nur «mechanisch» irgendwelche Schlüsselworte aufgreift und zu Sätzen eigener Fabrikation verwendet. Er soll auch dann sinnvolle Antworten geben, wenn das, wonach gefragt wird, in der Geschichte gar nicht ausdrücklich vorkam. Die Geschichte geht zum Beispiel so: «Ein Mann betritt ein Schnellrestaurant, bestellt einen Hamburger und verläßt das Lokal ohne zu zahlen.» Frage: «Hat er den Hamburger gegessen?» SAM: «Hat er nicht.» Er kann diese Antwort geben, weil er ein «Skript» normaler Restaurantsituationen besitzt, mit

dem er die Geschichte vergleicht. Es sieht, vereinfacht, die Folge Bestellen-Essen-Zahlen vor; wird nicht gezahlt, so «folgert» er, dann wird auch nicht gegessen worden sein. SAM scheint den nur impliziten Sinn der Geschichte verstanden zu haben. (Und möglicherweise ziehen auch Menschen Schlüsse dieser Art, weil sie «Skripte» vieler Normalsituationen parat haben und bei Bedarf heranziehen.)

Searles Test, mit dem er den «starken» Anspruch der AI ein für allemal niedergeschmettert zu haben glaubt, ist eine imaginäre Variante von SAM: ein Computer, der Chinesisch versteht. Und Searle selber spielt ihn.

«Nehmen Sie eine Sprache, die Sie nicht verstehen. Ich persönlich verstehe kein Chinesisch; für mich sind chinesische Schriftzeichen nur sinnlose Krakel. Stellen Sie sich nun vor, ich würde in ein Zimmer gesetzt, das Körbe voller Kärtchen mit chinesischen Symbolen enthält. Nehmen wir ferner an, man hätte mir ein Buch in meiner Muttersprache Englisch in die Hand gedrückt, das angibt, nach welchen Regeln chinesische Zeichen miteinander kombiniert werden. Dabei werden die Symbole nur anhand ihrer Form identifiziert, ohne daß man irgendeines verstehen müßte. Eine Regel könnte also sagen: 'Nimm ein Krakel-Krakel-Zeichen aus dem Korb Nummer 1 und lege es neben ein Schnörkel-Schnörkel-Zeichen aus dem Korb Nummer 2.' Angenommen, von außerhalb des Zimmers würden mir Menschen, die Chinesisch verstehen, kleine Stöße von Kärtchen mit Symbolen hereinreichen, die ich nach den Regeln aus dem Buch manipulierte; als Ergebnis reiche ich dann meinerseits kleine Kartenstöße hinaus. In die Computersprache übersetzt wäre also das Regelbuch das Computerprogramm, sein Autor der Programmierer und ich der Computer; die Körbe voller Symbole wären die Daten, die kleinen mir ausgehändigten Stöße die Fragen und die von mir hinausgereichten Stöße die Antworten. Nehmen wir nun an, das Regelbuch sei so verfaßt, daß meine Antworten auf die Fragen von denen eines gebürtigen Chinesen nicht zu unterscheiden sind. Beispielsweise könnten mir die Leute draußen eine Handvoll Symbole hereinreichen, die – ohne daß ich das weiß – bedeuten: 'Welches ist Ihre Lieblingsfarbe?' Nach Durcharbeiten der Regeln würde ich dann einen Stoß Symbole zurückgeben, die – was ich ebensowenig weiß – beispielsweise heißen: 'Meine Lieblingsfarbe ist blau, aber grün mag ich auch sehr.' Also hätte ich den Turing-Test für Chinesisch bestanden. Gleichwohl habe ich nicht die geringste Ahnung von dieser Sprache. Und ich hätte auch keine Chance, in dem beschriebenen System Chinesisch zu lernen, weil es mir keine Möglichkeit bietet, die Bedeutung irgendeines Symbols in Erfahrung zu bringen. Wie ein Computer hantiere ich mit Symbolen, aber verbinde keine Bedeutung mit ihnen. Der Punkt des Gedankenexperiments ist der: Wenn ich kein Chinesisch verstehe, indem ich lediglich ein Computerprogramm zum Verstehen von Chinesisch ausführe, dann tut das auch kein Digitalcomputer ... Das bloße Hantieren mit Symbolen genügt nicht für Fähigkeiten wie Einsicht, Wahrnehmung, Verständnis oder Denken. Und da Computer ihrem Wesen nach Geräte zur Manipulation von Symbolen

sind, erfüllt das bloße Ausführen eines Computerprogramms auch nicht die Voraussetzungen einer geistigen Tätigkeit.»

Nebenbei: Als Computer dieser Bauart wäre Searle ein Versager. In seinem Chinesischen Zimmer könnte er zwar grammatikalisch richtige Antworten geben, aber eine sinnvolle Antwort wäre höchstens durch Zufall einmal darunter. Denn das Regelbuch, nach dem er die chinesischen Schriftzeichen manipuliert, ist eine Syntax. Sie besagt also lediglich, welche Klasse von Zeichen vor und hinter welchen anderen Klassen von Zeichen stehen darf. Seine Antworten überzeugten niemanden, denn sie klangen eher wie «Meine Farbe ist schief, aber ärmlich kannst du auch stark». Um nicht nur richtige Zeichenklassen, sondern aus jeder Klasse das richtige einzelne Zeichen zu wählen, brauchte er ein Regelbuch, das mehr enthielte als eine Syntax, nämlich eine Semantik. Im Grunde aber änderte das an seinem Argument nichts. Denn die semantischen «Marken», die die Wörter in einem Computerlexikon tragen («belebt», «menschlich» und so weiter) sind ihrerseits abstrakte Symbole, die in sich keinen Verweis auf irgend etwas außerhalb ihrer selbst tragen. Sähe aber der Mensch im Chinesischen Zimmer all diese Marken, und lernte er es, geläufig mit ihnen umzugehen, so könnte er es kaum vermeiden, langsam auch die Bedeutung der Schriftsymbole zu erraten.

Searles Kritik zielt sehr viel tiefer als die seines Kollegen Hubert Dreyfus, des anderen beharrlichen Feindes vom Dienst. Dreyfus hält der AI im wesentlichen entgegen, daß es mit der Intelligenz heutiger Computerprogramme entgegen manchen Versprechungen noch nicht sehr weit her sei. Searle dagegen behauptet, menschliche Intelligenz sei überhaupt etwas anderes als ein Computerprogramm. Er schließt nicht aus, daß es eines Tages eine Maschine geben könnte, die sie besitzt. Er sagt aber: Eine Maschine, die nur das tut, was heutige Computer tun, nämlich nach einer Syntax mit (unverstandenen) Symbolen hantieren, wird sie mit Sicherheit nie besitzen.

Die wenigsten Verteidiger der AI in ihrer starken Fassung haben das intuitiv sofort Einleuchtende abzustreiten versucht: daß Searle in der Rolle des Computers zwar den Eindruck erzeuge, er verstehe Chinesisch, tatsächlich aber kein Wort versteht. (Nur wenige meinten, auf eine ihm selber nicht einsichtige Weise verstehe er es im Chinesischen Zimmer vielleicht doch.) Die meisten Einwände liefen auf etwas anderes hinaus. Sie variierten den Gedanken, daß Searles auf Anhieb so einleuchtende Allegorie die Ebenen verwechselt habe. Gewiß, er erwecke den Anschein, Chinesisch zu können, verstehe aber gar keins — die Neuronen im Schädel eines echten Chinesen verstünden es indessen genauso wenig. Der ganze Chinese sei es, der Chinesisch könne — und so könne man auch von dem ganzen Chinesischen Zimmer sagen, es verstehe Chinesisch. Verstehen sei ein Attribut des ganzen Systems, nicht seiner mechanischen Einzelteile.

Worauf Searle erwidert: «Stellen Sie sich vor, ich hätte den Inhalt der Körbe und das Regelwerk auswendig gelernt und würde alle Berechnungen in meinem Kopf durchführen; Sie dürfen sich sogar vorstellen, daß ich unter freiem Himmel arbeite. Es gibt dann nichts an diesem 'System', das sich nicht in mir befände – und da ich kein Chinesisch verstehe, versteht es auch das System nicht.»

Eine andere Gruppe beliebter Einwände stößt sich sozusagen an den Größenverhältnissen in Searles Allegorie: Das Chinesische Zimmer mit seinem einen mit Symbolen hantierenden Element, Mr. Searle, sei keine Abbildung des Gehirns, in dem zehn Milliarden Neuronen Signale miteinander tauschen. Dem setzt Searle seine Chinesische Turnhalle entgegen – eine ganze Schar von Leuten, die mit chinesischen Symbolen hantieren, ohne Chinesisch zu verstehen. Eine solche Turnhalle, sagt Searle, verstehe genauso wenig Chinesisch wie das Zimmer. Sie sei, so sagt die andere Seite, aber auch ebenfalls keine angemessene Abbildung des Gehirns. Wenn nämlich jedes Hirnneuron von einer Person gespielt werden soll, müßte sie die menschliche Bevölkerung von mehr als zehntausend Erden fassen – und das ergäbe möglicherweise dann doch ein denkendes Gehirn, «monströs, langsam, merkwürdig konstruiert, aber dennoch vielleicht funktionstüchtig» (die Philosophen Paul und Patricia Churchland).

Der Einwand entspricht einer Stimmung, die in den Kreisen der AI und der Kognitionswissenschaft heute weit verbreitet ist: Höhere geistige Funktionen seien eine Folge schierer Größe – man mache ein System nur komplex genug, und zwangsläufig müßten sie sich alle miteinander einstellen, müßten «emergieren» (wörtlich: auftauchen). Solcher «Emergentismus» schimmert auch bei Douglas Hofstadter durch. «Die Tatsache, daß Intelligenz, Verstand, Geist, Bewußtsein, Seele und all das einer unwahrscheinlichen Quelle entstammen – einem enormen Gewirr von Zellkörpern, Axonen, Synapsen und Dendriten –, ist absurd und dennoch unbestreitbar. Wie dergleichen ein 'Ich' hervorbringen kann, ist schwer zu verstehen, aber wenn wir diese fundamentale, seltsame, beunruhigende Tatsache einmal akzeptiert haben, dann sollte es nicht sonderbarer sein, ein 'Ich' aus Wasserrohren zu akzeptieren.» Eine verwegene Zuversicht. Komplexität ist sicher eine notwendige Vorbedingung für höhere geistige Funktionen; aber daß sie auch eine ausreichende sei, das ist keineswegs ausgemacht und eher unwahrscheinlich. Kein Einzelteil eines Radioempfängers hat alle Eigenschaften eines Radios, nur in geringerem Maße. Ein Radio ist nicht die Summe der Eigenschaften seiner Komponenten. Als Radio «emergiert» es erst, wenn die Einzelteile zu einem ganz speziellen System zusammengeschaltet sind. Es nützte nicht das mindeste, sie alle irgendwie zusammenzuwerfen und, wenn die Sache noch nicht funktioniert, weitere Komponenten hinzuzutun und das Gebilde immer komplexer zu machen. So etwas wie Rundfunkempfang ließe sich durch ein wenig komplexes System wahrscheinlich nicht realisieren; aber Komplexität selber hülfe noch gar nicht – auf das Wie käme es an. Und auch das al-

lerenormste Gewirr von Zellkörpern, Axonen, Synapsen und Dendriten machte selbst noch kein verständiges Gehirn.

An Searles Argument geht der Einwand, seine Denkmodelle seien einfach nicht komplex genug, völlig vorbei. Er sagt nämlich sehr deutlich, was seiner Meinung nach den Computer vom denkenden Menschen unterscheidet: Es fehlt seinen Operationen die Intensionalität. Das heute vielstrapazierte Wort ist wenig glücklich, denn es stiftet Verwirrung und beschwört endlose Mißverständnisse herauf. Mit den «Intentionen», den Absichten der Umgangssprache und der Volkspsychologie hat es jedenfalls nichts zu tun. Aber die «*intentionality*» der heutigen angelsächsischen Philosophie ist auch nicht völlig identisch mit dem, was in der Logik «Intension» heißt: der Begriffsinhalt eines Ausdrucks, sein Sinn, das, was er uns zu verstehen gibt – im Unterschied zu seiner «Extension», die nur angibt, welchen Gegenständen er zukommt, also seinen Bedeutungsumfang. (Die beiden Aussagen «Der Abendstern steht am Himmel» und «Der Morgenstern steht am Himmel» sind extensional gleichbedeutend, denn es sind gleiche Aussagen über den gleichen Gegenstand, den Planeten Venus. Der Unterschied besteht allein in der Intension der Begriffe ‘Abendstern’ und ‘Morgenstern’, das, was wir ihnen als ihren Sinn entnehmen.) Trotzdem schimmern bei vielen Erörterungen der «*intentionality*» die Bedeutungen ‘Absicht’ oder ‘Begriffsinhalt’ hindurch.

Was also ist «Intensionalität»? Mag einer der Hauptexponenten der heutigen Debatte sie selber definieren, der amerikanische Philosoph Daniel Dennett: «Intensionalität ist Bezogenheit (*aboutness*). Manche Dinge sind auf andere Dinge bezogen: Eine Meinung kann man *über* Eisberge haben, aber ein Eisberg ist nicht *über* etwas anderes, ist auf nichts anderes bezogen; ... ein Buch oder Film kann über Paris sein, aber Paris handelt von nichts.»

Intensional in diesem Sinn also heißt schlicht: auf etwas außerhalb seiner selbst bezogen, auf etwas anderes «hingerichtet» (mit dem Wort des Philosophen Franz Brentano, der in dieser Intensionalität das Kennzeichen alles Geistigen sah). Das Gegenteil wäre etwa «formal». Etwas sei «formal richtig» heißt: Es ist logisch in sich konsistent, ohne Rücksicht darauf, ob es eine Entsprechung zu etwas außerhalb hat.

Was tut der Computer? Nach außen wirkt er, als verarbeitete er auf vielfältige Weise die ihm eingegebenen «Daten»; im Innersten manipuliert er Zahlen, indem er sie den Grundrechenarten und einer kleinen Handvoll von Operationen der Prädikatenlogik unterwirft («und», «oder», «entweder-oder», «nicht» sowie das Boolesche «wahr/falsch»); und noch tiefer drinnen ist er ein Schaltkreis, der elektrische Potentiale verteilt. Die Zahlen sind Symbole: für Zahlen, für Buchstaben, für was auch immer der Programmierer wollte. Welche Operationen er in welcher Reihenfolge an ihnen ausführt, schreibt ihm das «Programm» vor. Das Programm ist eine Liste aufeinander bezogener Befehle: eine «Syntax». Eine Syntax ist eine Sammlung formaler Regeln. Sie hat nichts Intensionales, ist auf nichts

außer sich selber bezogen. Genauer besehen, sind nicht nur die Symbole, an denen der Computer sich zu schaffen macht, für ihn selber gar keine, sondern nur für seinen Benutzer; auch die Syntax wird ihm erst eingeflüstert. «So wie der Syntax keine Semantik innewohnt, so wohnt der Physik keine Syntax inne.» Beides, Semantik wie Syntax, wird der Maschine erst vom Menschen zugewiesen. Der menschliche Geist, sagt Searle, habe aber zur Syntax auch eine Semantik: Die Symbole, mit denen er hantiert, beziehen sich auf etwas außerhalb ihrer selbst, bedeuten etwas. Was ihn vom Computer trennt, sei also die Intensionalität seines Geistes. Nie werde die bloß formale Syntax eines Computerprogramms Geist ergeben.

So formuliert, leuchtet Searles These aber gleich sehr viel weniger ein. Auf der untersten Signalebene hantiert der Computer mit elektrischen Ladungen, das Gehirn mit den Entladungsfrequenzen seiner Neuronen. Auf dieser Ebene «wissen» beide nicht, daß es sich um Symbole für irgend etwas handelt, daß sie eine Syntax exekutieren, ganz zu schweigen von irgendeiner Semantik. Irgend etwas im Gehirn weist ihnen Symbolwert zu. Aber warum sollte das dem Computer unmöglich sein? Versteht man «intensional» strikt als «über» oder «auf etwas anderes bezogen», so handeln auch die endlosen Einsen und Nullen des Computers sehr wohl von etwas außerhalb ihrer selbst. Nur ist der Computer nicht von selber drauf gekommen; die Zuweisungen nehmen seine Programmierer und Benutzer vor. In diesem Sinn besitzt der Computer durchaus auch Intensionalität.

Wer sagt, die Intensionalität des menschlichen Verstandes sei jedenfalls kein von außen eingespeistes Programm, sondern stelle sich nur darum ein, weil sich der ganze Mensch ständig aktiv mit seiner Umwelt auseinandersetze, dem tritt sofort ein einstweilen nur virtueller Roboter entgegen. Er ist eine denkbare Maschine, die sich ebenfalls selber aktiv mit ihrer Umwelt auseinandersetzt: mit allerlei Technik bestückt, die Sinnesorgane imitiert, mit «Augen» und «Ohren» und Tastsensoren und Temperaturfühlern, und so programmiert, daß er die einlaufenden Daten in irgendein Verhalten umsetzt und dabei aus Erfolgen und Mißerfolgen lernt. Schlechterdings wäre auch er ein intensionales System, das außerdem noch auf eine dem Menschen ähnliche Art zu seiner Intensionalität gekommen wäre.

Begriffe wie «Geist» und «Verstand» sind undefinierbar. Jeder ist frei, etwas anderes darunter zu verstehen, und so reden meist alle baby-lonisch aneinander vorbei. Ein gewisser Zugewinn an Präzision stellt sich ein, sobald man statt dessen bestimmte Fähigkeiten ins Auge faßt, durch die der ungreifbare «Geist» sich auszeichnet: Er nimmt wahr, er denkt, er versteht, er sieht ein, er fühlt, er meint, er verfolgt Ziele, er lenkt intelligent das Verhalten. Davon läßt sich dann schon eher ohne Verwirrung reden: von Wahrnehmungen, Gefühlen, Absichten, Schlußfolgerungen. Es sind alles intensionale Fähigkeiten. «Denken», «meinen», «fühlen» kann man nicht schlechthin; immer denkt man, daß (der Turm schief ist), meint man, daß (das Wetter morgen besser wird), fühlt man, daß (der Kopf wehtut).

Daß der Computer viele dieser Fähigkeiten hat oder jedenfalls eines Tages haben könnte, ist gar keine Frage mehr. Er verarbeitet von außen kommende Daten, nimmt also wahr. Wenn ich seine Schachzüge ersinnen müßte, müßte ich denken; und wenn er mich schlägt, hat er besser gedacht. Eine computergesteuerte Rakete verfolgt allerdings ein Ziel. Ein Autopilot steuert das Verhalten des Flugzeugs, und er simuliert das Fliegen nicht nur, sondern fliegt es «in echt», und wir vertrauen ihm unser Leben an. Selbst Gefühle ließen sich ihm beibringen. Wenn er einerseits voraussieht, daß eine Berechnung lange dauern wird, andererseits registriert, daß sein Benutzer mehrmals ungeduldig auf die Eingabetaste schlägt, könnte er den Bildschirm rosa einfärben und rasch wirre Zahlen herunterhaspeln — so als schämte er sich und wäre nervös. (Wohlgemerkt, es handelte sich um gefühlsmäßiges *Verhalten*; daß er dabei selber etwas empfindet, ist nicht gesagt, und wer würde es glauben?)

Solange man in behavioristischer Tradition nur auf den Output sieht, dürfte es wohl keine Fähigkeit geben, die ihm grundsätzlich verwehrt ist. Er erarbeitet, wozu der Mensch seinen Verstand braucht; er ist eine «ratiomorphe» Maschine. Und da der Begriff Intelligenz sich mehr an äußeren Leistungen orientiert als an den inneren Prozessen, aus denen diese hervorgehen, wird man Computern eines Tages auch das Attribut intelligent sicher nicht mehr absprechen wollen.

2

Gehirn und Computer: oft ist hervorgehoben worden, was sie gemeinsam haben – und als Reaktion darauf dann auch, was sie trennt. Sind es Abgründe?

Auf die Gemeinsamkeiten gründet sich der produktivste Zweig der heutigen Psychologie, die Kognitionspsychologie, die untersucht, wie «Wissen» in den Kopf kommt, wie es dort gespeichert, verwaltet, abgerufen wird – die Welt als Repräsentation im Zentralnervensystem, das Geistorgan als informationsverarbeitende Maschine. Und so, wie man das Gehirn als eine Art biologischen Computer sieht, sieht man den Computer gern als eine Art maschinelles Gehirn, ein «Elektronengehirn» eben.

Zwar hat es nicht den Anschein, als käme die Kognitionspsychologie ins Stocken oder als ginge ihr in absehbarer Zukunft die Arbeit aus; aber ob sie das Gehirn restlos als eine Art von Computer erklären kann, steht einstweilen völlig dahin. Auffällig, geradezu verdächtig wenig interessiert sie sich bisher für subjektive Erlebnisqualitäten — für das Eigentümliche eines Schmerzgefühls etwa oder für Wahrnehmungsqualitäten wie «grün» oder «heiß». Rein funktionalistisch mag sie eines Tages beschreiben können, wie das Gehirn eine Farbe wie «grün» repräsentiert und welchen Zwecken es dient, überhaupt Farben unterscheiden zu können; aber wie «grün» aussieht, ja auch nur, ob das «Grün» des einen tatsächlich wie das «Grün» des anderen oder vielmehr wie dessen «Rot» oder auch völlig an-

ders aussieht — das läßt sie auf sich beruhen. Funktional gleich nämlich wäre auch ein umgekehrtes Spektrum: Es ließe exakt die gleichen Unterscheidungen zu, erfüllte exakt den gleichen Zweck; und aus der Aktivität bestimmter Neuronenbahnen in der Sehrinde läßt sich nicht im mindesten schließen, wie etwa eine Farbe subjektiv erlebt wird.

Welche sind die offenkundigen Gemeinsamkeiten von Gehirn und Computer? Wenn manche Begriffe auf beide zutreffen, dann nicht nur in einem ungefähren metaphorischen Sinn, sondern weil sie tatsächlich Gleiches beim Namen nennen.

Beide gehen mit «Wissen» um: erhalten von außen verschiedenerlei Nachrichten über Sachverhalte (über Gegenstände oder Vorgänge oder Beziehungen), das Gehirn mittels der Sinnesorgane, der Computer mittels Tastatur, Scanner, Mikrophon, und «verarbeiten» diese «Information»: speichern sie, manipulieren sie, indem sie sie ordnen, nach bestimmten Regeln verändern, Schlüsse aus ihr ziehen, geben sie bearbeitet auf verschiedene Weise wieder heraus und können damit «Verhalten» steuern (die Bewegungen einer schreibenden Hand, die Hitzeimpulse der Tinten-düsen eines Druckers).

Beide besitzen mit diesen Informationen ein inneres «Bild», eine innere Repräsentation bestimmter Sachverhalte der äußeren Welt.

Beide sind lernfähig: imstande, neue Informationen aufzunehmen und ältere in ihrem Licht zu revidieren.

Beide sind imstande, Ziele zu verfolgen.

Beide «kodieren» die Informationen, mit denen sie hantieren, auf eine spezifische Weise, und zwar eine Weise, die keinerlei Ähnlichkeit mit dem Abgebildeten hat: Im Gehirn befindet sich kein sichtbares Abbild dessen, was die Augen gerade sehen, es riecht in ihm auch nicht, wenn es gerade einen Ammoniakgeruch wahrnimmt — so wie die Bitmap, die der Computer von einer gescannten Seite anlegt, selber kein sichtbares Abbild dieser Seite ist, sondern kodierte Information darüber, wie diese Seite aussieht, also eine Folge von abstrakten Zeichen, von Signalen.

Beide, Gehirn wie Computer, arbeiten mit elektrischen Signalen.

Beide haben die Fähigkeit der zeitlichen Integration: Sie besitzen ein «Gedächtnis», das frühere Informationen verwahrt, vergleichen den aktuellen Input mit dem Inhalt dieses «Speichers» und können aus diesen Vergleichen Schlüsse auf Zukünftiges ziehen.

Und welches sind die offenkundigen Unterschiede?

Das Gehirn mit seinen zehn Milliarden Nervenzellen, von denen jede Signale mit bis zu zehntausend anderen austauscht, ist um Größenordnungen komplexer als selbst der größte heutige Computer.

Die elektrischen Signale werden im Gehirn nicht wie im Computer durch Leitung weitergegeben, sondern auf elektrochemischem Weg: in-

dem an den Kontaktstellen zwischen den Neuronen Chemikalien (Neurotransmitter) ausgeschüttet werden, die im anderen Neuron elektrische Potentiale erzeugen.

Informationen kodieren beide auf verschiedene Weise: der Computer letztlich durch Sequenzen von 0 und 1, dargestellt durch Sequenzen von elektrischer Ladung und Nicht-Ladung; das Gehirn durch die Zeitintervalle zwischen den uniformen Signalen seiner Neuronen, also durch die Frequenzen seiner Signale.

Der normale Computer hat einen einzigen Prozessor für die Manipulation der Signale oder höchstens ein paar; im Gehirn ist jedes Neuron ein spezialisierter Prozessor, der die ankommenden Signale summiert (und die hemmenden subtrahiert), um selber aktiv zu werden, wenn eine bestimmte Summe erreicht ist.

Infolgedessen arbeitet das Gehirn «parallel», der Computer «seriell». Das heißt, im Gehirn finden Millionen einzelner Signaltransformationen gleichzeitig statt, im gewöhnlichen Computer immer nur eine nach der anderen. Dafür ist die Signalübermittlung im Computer millionenmal schneller. Wo es um die monotone Wiederholung ein und derselben Operation geht, die auch das Gehirn nur seriell abarbeiten kann (also etwa von eins bis hunderttausend hochzuzählen und dabei jeweils drei zu überspringen), ist darum der Computer ungleich fixer; wo jedoch parallele Techniken zum Zuge kommen können (etwa wenn ein Bild auf ein paar relevante Merkmale zu analysieren ist), ist das Gehirn weit überlegen.

Das Gehirn besitzt, was heutige Computer nicht haben: die Fähigkeit der Selbstorganisation — es entwirft sich aus dem Erfolg oder Mißerfolg früherer «Programme» neue eigene Handlungsstrategien.

Der Computer kann eine Reihe logischer Operationen an Symbolen vornehmen, weiß aber «von Hause» aus nicht, welche Informationen er wie zu bearbeiten hat; dies teilt ihm erst das jeweils aktive Programm mit. Das Gehirn ist «von Natur» aus zu einem erheblichen Teil vorprogrammiert, «weiß» also etwa von vornherein, wie es Farbinformationen aus den Botschaften der Augen extrahiert und Klanginformationen aus den Bewegungen der Zilien im Innenohr. Diese Vorprogrammierung besteht in der topologischen Gestalt seiner Neuronennetze, also in den Kontakten, die zwischen spezifischen Zellen eingerichtet sind. In den Wachstumsjahren knüpfen die Nervenzellen ihre Verbindungen: Chemische Lockstoffe lassen die Fortsätze in bestimmte Zielgebiete hineinwachsen; aber nur jene Kontakte, die dann wirklich benutzt und durch die Benutzung stabilisiert werden, bleiben bestehen.

Der Computer tritt immer nur dann in Aktion, wenn er dazu eigens aufgefordert wird, das Gehirn auch von sich aus, «spontan».

Das Gehirn hat ein biologisches Schicksal: Es wird geboren, wächst, reift, sammelt Information, erkrankt möglicherweise, altert und stirbt. Sein Zweck bei allem ist es, ein biologisches Wesen sicher durch das Le-

ben zu lotsen. Der Computer hat das Schicksal einer von einem biologischen Gehirn gebauten Maschine; sein Zweck ist es, genau das zu tun, was biologische Gehirne ihm auftragen.

Ob die Ähnlichkeiten oder die Unterschiede größeres Gewicht haben, läßt sich schwer sagen. Falls sich unter den Unterschieden, den genannten wie den ungenannten und vielleicht bis heute unbekanntem, nicht einer findet, der einen unüberwindbaren kategoriellen Unterschied bedeutet, möchte ich meinen: die Ähnlichkeiten. Jedoch dürfen die Unterschiede nie aus dem Auge verloren werden. In einzelnen Fall, bei jeder konkreten Frage (etwa: wie werden Erinnerungen gespeichert, wie wieder aufgefunden?) könnten sie die beiden informationsverarbeitenden Systeme nämlich dann doch unvergleichbar machen.

Daß sie aber überhaupt je in irgendeiner Hinsicht vergleichbar sind, daß gewisse geistige Aktivitäten auf dem Computer simuliert werden können und daß das, was der Computer tut, sich mit mentalistischen Begriffen beschreiben lässt — das hängt sozusagen vom Ausgang einer Wette ab, die eine Mehrzahl heutiger Wissenschaftler gegen die geballte abendländische Tradition abgeschlossen haben: der Wette, daß geistige Phänomene eine Hervorbringung der materiellen, der physikalischen Welt sind und nichts ganz Andersartiges, auf das deren Gesetze nicht zutreffen, der Wette, daß Geist keiner anderen Welt zugehörig ist als die ganze übrige Natur. Es ist die monistische Wette.

Das ganze westliche Denken aber war immer zutiefst dualistisch und ist es bis heute. Hier der Körper, dort der Geist; hier der Leib, dort die Seele. Dualismus heißt in dieser Sache: zu meinen, daß geistig-seelische Phänomene (also das, was die Volkspsychologie mit Begriffen wie Verstand, Vernunft, Denken, Wahrnehmung, Erinnerung, Gefühl, Empfindung, Wille, Bewußtsein, Motivation und so fort belegt) eine autonome, von der materiellen Welt unabhängige Wirklichkeit besitzen, auf die die Naturgesetze nicht zutreffen und die mit ihrer Hilfe folglich auch nie erklärt werden kann.

Ein großer Dualist stand am Anfang der Philosophie der Neuzeit: René Descartes, der die ausgedehnte Welt der Materie (die *res extensa*) von der ausdehnungslosen, unkörperlichen Welt des mathematischen und sprachlichen menschlichen Denkens schied (der *res cogitans*). Auch er schon stand vor einer unbequemen Frage, der sich kein rational denkender Dualist entziehen kann: Wie und wo verkehren die beiden so grundverschiedenen *res* miteinander? Ihretwegen erfand sich Descartes «animalische Geister», die von der Zirbeldrüse aus (einer Gehirnstruktur, die ihm wegen ihrer Unpaarigkeit aufgefallen war) durch den Körper geschwemmt werden und überall in ihm auf die ausgedehnte Materie einwirken sollen.

Descartes' dualistische Position hatte Folgen, die bis heute nicht abgetragen sind. Da Tiere nicht rechnen und sprechen, verwies er sie in die dumpfe Welt der Materie; auf ihn berief sich seither (wenn auch nicht ganz zu Recht) die Ansicht, Tiere seien bloße Automaten, die man dann

auch als solche behandeln dürfe — die Eiermaschinen in ihren Batterien wüßten ihm wenig Dank.

Ein Erzdualist aber war natürlich auch schon Plato, als er in die abendländische Tradition die mystische orphische Lehre einspeiste, die Seele sei etwas Besseres als der Körper, etwas Immaterielles und Unvergängliches, das beim Tode aus dem körperlichen Gefängnis befreit werde.

Im Grunde aber ist die Fahndung nach bestimmten Urhebern des Leib-Seele-Dualismus müßig. Alle Unsterblichkeitsreligionen setzen ihn voraus und haben bis heute Interesse an seinem Wohlergehen. Und schon in der engen, unverständenen und bedrohlichen — also wahrhaft unheimlichen — Welt des vorgeschichtlichen Menschen dürften körperlose Geister in großer Zahl gespukt haben. Etwas wurde bewirkt, und man wußte nicht, wodurch: Eine Wolke krachte, ein Fluß versiegte, ein Berg spie Feuer — da konnten nur unkörperliche, weil unsichtbare Wesen am Werk gewesen sein.

Aber auch ohne jede ideologische Beihilfe wären wir vermutlich alle Dualisten. Unsere eigene innere Erfahrung — unsere Introspektion — machte uns dazu. Wir erleben nicht nur unseren Geist als geisterhaft: als etwas, das keine Ausdehnung und keinen Ort und keine Masse hat – es führt in unserer Innenschau auch kein Weg zu etwaigen körperlichen Substraten des Denkens. Selbst daß es das Gehirn ist, welches denkt, wissen wir nicht aus eigenem Erleben, und frühere Zeiten haben ganz andere Organe für den Sitz des Denkens gehalten. Überall in der Natur können wir dadurch, daß wir dichter herangehen, mehr erkennen, auch Dinge, deren Existenz der distanziertere Blick nicht einmal ahnen ließ. Von weitem ist ein Blatt nur eine grüne Fläche, aus größerer Nähe bemerken wir seine Strukturen, mit einem Mikroskop erschließen sich seine Zellen, mit einem Elektronenmikroskop deren Membranen und die größeren Moleküle, und dieser Zoom ist so wunderbar aufschlußreich und schlüssig, daß wir keine Schwierigkeit mehr haben, den Physikern zu glauben, daß die Moleküle aus Atomen und diese aus Elementarteilchen bestehen. Unsere Introspektion aber können wir anstrengen, soviel wir wollen — sie erkennt kein Nervengewebe, keine Neuronen, keine elektrische Entladungen an den Synapsen, obwohl doch auch die entsprechende geistige Anstrengung selber mit einer solchen Aktivität einhergeht. Unser bewußtes Erleben reicht nicht bis zu seinen Wurzeln hinab. Es bleibt an seine eigene Ebene gefesselt.

Da wir unser Denken und Fühlen als etwas Immaterielles erleben, neigen wir dazu, es auch für einen immateriellen Vorgang zu halten. Und selbst eingefleischte Monisten kommen nicht umhin, von geistigen Prozessen in dualistischen Begriffen zu sprechen, denn es gibt gar keine anderen. Die Sprache ist durchtränkt von Dualismus. Da sind nicht nur feierliche poetische Sinnsprüche wie «Stark am Geist, am Leibe schwach», die Dualismus predigen. Die Alltagssprache tut es immerfort: «Seine Angst trieb ihn um», «Sie tanzte vor Freude», «Er ekelte sich so, daß er sich ü-

bergeben mußte» – immer ist es etwas Psychisches, das dem Körper Beine macht. Wir erleben die Psyche nicht als einen Bestandteil des körperlichen Lebens, sondern als etwas, das ihm gegenübersteht und das sie beherrscht oder doch beherrschen sollte, denn als guten Platonikern und Christen erscheint uns die Psyche allemal auch vornehmer als der niedere Körper, und so drücken wir uns denn auch aus.

So verführerisch und geradezu unentrinnbar aber das Argument der Introspektion auch zu sein scheint, stichhaltig ist es nicht. Auf allen Gebieten haben wir zur Kenntnis nehmen müssen, daß man seinen Augen nicht immer trauen darf, daß der naive Blick nicht unbedingt auch richtige Erkenntnisse liefert. Die Erde war denn doch keine Scheibe, über der sich ein Himmelsgewölbe drehte; nicht Gestank rief Krankheiten hervor, sondern Mikroben, die mit bloßem Auge bei bestem Willen nicht zu auszumachen waren. Hinter der Wirklichkeit, die wir ohne Hilfsmittel wahrnehmen, stecken allerorten andere, unglaublich reiche und zuweilen geradezu konterintuitive Wirklichkeiten. In dem Maße, in dem wir dahinter kamen, ist auch klar geworden, daß uns unsere Wahrnehmungen keineswegs ein Bild der Welt als solche liefern, der Welt «an sich». Wie sie an sich beschaffen ist, könnten wir unmöglich sagen; wir erleben immer nur eine artspezifisch interpretierte Wirklichkeit. Wir sind nämlich ganz auf die uns genetisch einprogrammierten Anschauungsformen unserer paar Sinne angewiesen, die schnelle Luftdruckschwankungen zu «Tönen» machen, die Anwesenheit bestimmter Aerosole zu «Gerüchen», die reflektierte elektromagnetische Strahlung mit einer Wellenlänge von 500 bis 550 Nanometer zur «Farbe grün». Die Evolution hat uns mit ein paar Wahrnehmungen ausgestattet, die es unseren Ahnen ermöglicht haben, sich erfolgreich in der Biosphäre dieses Planeten zu bewegen und zu behaupten; ein direkter Zugang zur Wirklichkeit der Atome und Galaxien, zu Mikrokosmos und Makrokosmos war nicht nötig und ist uns darum auch nicht gegeben. Und es reichte in all den Jahrtausenden, daß wir eine Furcht *empfanden*, eine Kontur *sahen*, einen Schmerz *fühlten*, eine Bewegung ausführen *wollten*. Zu verfolgen, wie das alles zustande kam, war nicht überlebenswichtig, und so entzieht es sich offenbar der unmittelbaren Einsicht und läßt sich nur mittelbar mit den Methoden der Wissenschaft erschließen wie irgend etwas völlig Fremdes.

Unserer Introspektion sind also natürliche Grenzen gesetzt. Auch was sie niemals bestätigt, kann sehr wohl der Fall sein.

Was aber spricht gegen den Dualismus? Was spricht dafür, daß mentale Phänomene nicht zu einer separaten Welt gehören?

Jeder Dualismus verträgt sich schlecht mit unserm von Tag zu Tag wachsenden Wissen über die tatsächlichen Beziehungen, die zwischen dem Gehirn und dem psychischen Geschehen bestehen. Hirnverletzungen und operative Eingriffe haben immer wieder demonstriert, daß bestimmte Läsionen bestimmte geistige Funktionen verändern und beeinträchtigen. Daß Geistesranke von Dämonen besessen seien, wurde immer unglaub-

hafter; heute ist klar, daß Geisteskrankheiten hirnorganische Funktionsstörungen sind. Mit einer Vielzahl von chemischen Substanzen lassen sich geistig-seelische Vorgänge dämpfen, verstärken, verändern, verzerren, ausschalten. Wenn der bewußte Geist etwas Andersartiges ist, letztlich nicht abhängig von irgendeinem materiellen Geschehen – warum verwirren ihn dann Lysergsäurediäthylamidmoleküle (gemeinhin als LSD bekannt) so schön und schrecklich? Warum läßt er sich von Diäthyläthermolekülen oder einem primitiven Knüppelhieb ganz in die Flucht schlagen?

Jeder Dualismus verstößt gegen die Annahme, daß es in der Natur durchweg mit natürlichen Dingen zugeht. Sie hat sich seit Anbeginn der Wissenschaftsgeschichte als wie man weiß ziemlich erfolgreich und bisher an keiner Stelle als ungültig erwiesen. Für vieles in der Welt wurden einmal über- oder außernatürliche Gründe angenommen, die sich unter dem wissenschaftlichen Blick in nichts auflösten. Blitze sind keine göttlichen Zornbekundungen, das Leben ist nicht Produkt einer übernatürlichen Lebenskraft. Da steht die Wette nicht schlecht, daß es auch im Schädelinneren keine zweite Welt neben oder in oder hinter der ersten gibt.

Letztlich muß jeder Dualismus der Erkenntnis widersprechen, daß das Universum ein geschlossenes System ist, in dem Energie weder verloren gehen noch neu entstehen kann. Wenn Geist – ein Wunsch, ein Wille, ein Entschluß – den Körper bewegen kann, dann sickerte an dieser Stelle Energie ins Universum ein; und wenn ein körperlicher Reiz vom Geist unterdrückt wird, dann würde Energie vernichtet. Dem wird nicht so sein. Das Universum wird keine Ritze, kein Leck haben, und schon gar nicht ausgerechnet im Kopf des Menschen.

Daß Geist viel, sehr viel mit Gehirnaktivität zu tun hat, können heute auch Dualisten kaum noch bestreiten. Sie stehen damit vor der Frage, wie das Immaterielle mit dem Materiellen interagiert. Der australische Hirnforscher und Nobelpreisträger John C. Eccles, zusammen mit dem Philosophen Karl Popper einer der letzten standhaften Dualisten, offeriert in seinem neuen Buch Die Evolution des Gehirns dazu eine wissenschaftliche Hypothese. Ein «Psychon» (wie er eine angeblich völlig autonom existierenden mentale Einheit nennt), sagt er, «durchdringe» jeweils ein «Dendron», nämlich ein Bündel von Dendriten in der Großhirnrinde (Dendriten sind die «Empfangsantennen» der Neuronen), und löse damit ein neurales Ereignis aus. Wie soll das möglich sein ohne Verstoß gegen das Energieerhaltungsgesetz? Weil das Psychon zu einem «Feld im herkömmlichen physikalischen Wortsinne» gehört, aber einem immateriellen Feld, das einem quantenmechanischen Wahrscheinlichkeitsfeld analog ist; im quantenphysikalischen Maßstab lasse sich Energie borgen. Physiker werden wahrscheinlich sofort abwehrend die Hände heben: Das aller kleinste wirkungsvolle neurale Ereignis, die Öffnung eines einzigen Neurotransmitterbläschens am Ende eines Axons, ist immer noch eine Sache von etwa zehntausend Molekülen, und in diesem Maßstab gelten keine quantenphysikalischen Begründungen. Laien werden jede Aussage über die Existenzform

dieses angeblich autonomen Geistes vermissen (außer daß er ein «Geschenk Gottes» sei). Zur Ontologie: Wo kommt er her? Wo geht er hin? Wie ist er in sich beschaffen? Zur Funktion: Warum gibt es ihn? Wie kommt er in den Kopf? Wie arbeitet er mit ihm zusammen? Zur Genese: Ist er wie der Organismus ein Produkt der Evolution? Reift und altert auch er? Da hat ein Theologe den Talar gegen einen weißen Kittel getauscht.

Aber wenn der Dualismus die schwierigen Fragen auch vervielfacht – der Monismus hat sie keineswegs gelöst.

Einige seiner Spielarten, die nicht minder zahlreich sind als die des Dualismus, sind schon im Ansatz völlig ungenügend, und es ist eine Freude, wie der kanadische Wissenschaftstheoretiker Mario Bunge (in seinem Buch *The Mind-Body Problem*, 1980) nicht nur mit den verschiedenen Dualismen, sondern auch mit ihnen umspringt, daß die Fetzen fliegen.

Der «eliminative Materialismus» in seiner radikalen Form bestreitet rundweg die Existenz aller mentalen Phänomene. «Er bietet keine Lösung für das Leib-Seele-Problem; nein, er behauptet, es gebe ein solches Problem gar nicht. Darum können wir den eliminativen Materialismus selber eliminieren.» In seiner mildesten Form besagt er nur, daß die Begriffe der Volkspsychologie (Zartgefühl, Willenskraft und so fort) unverlässlich sind und daß eine wissenschaftliche Beschreibung des Mentalen im Lichte der Hirnphysiologie möglicherweise zu ganz anderen begrifflichen Unterscheidungen käme. Das ist gut möglich. Vielleicht wissen wir eines Tages, daß es «die Angst» gar nicht gibt, sondern eine Reihe ganz verschiedener angstvoller Zustände; und auch, welche Maschinerie diese hervorbringt. Dann sagen wir vielleicht nicht mehr «Ich habe Angst», sondern «Ich befinde mich in Zustand 17c, gekennzeichnet durch Neurotransmitterabundanz Soundso im kaudalen Hypothalamus». Vielleicht aber auch lieber nicht.

Der Physikalismus, auch als reduktiver Materialismus bekannt, sagt: Geistige Zustände sind nichts anderes als physische Zustände des Gehirns, beides ist ein und dasselbe; wenn eines Tages das Gehirn vollständig beschrieben sei, werde alles Psychische automatisch miterklärt sein. Bunge: «Diese reduktionistische These ist falsch. Zu sagen, das Gehirn bestehe aus einer Menge Zellen, heißt nicht, das Gehirn sei nichts als eine Menge Zellen ... Ein System ist nicht dasselbe wie die Menge seiner Komponenten.»

Nein, sagt Bunge (und mich überzeugt er damit), auch eine materialistische Theorie darf das Geistige – also das, was sie erklären sollte – nicht einfach wegeskamotieren. Es bleibe nur eine Denkmöglichkeit: der «emergentistische Materialismus». «Er behauptet, daß das Zentralnervensystem weit davon entfernt ist, ein physikalisches Etwas – insbesondere eine Maschine – zu sein; vielmehr ist es ein Biosystem, das heißt ein komplexes Ding mit Eigenschaften und Gesetzen, wie sie nur lebenden Wesen eigen sind; sehr besonderen Eigenschaften und Gesetzen sogar, die nicht alle Biosysteme gleichermaßen besitzen. Die für das Geistige reklamierte

‘Emergenz’ ist eine doppelte: Die einzelnen Zellen besitzen die mentalen Eigenschaften eines Zentralnervensystems nicht; diese sind vielmehr systemische Eigenschaften, und darüberhinaus solche, die sich nicht von selbst ergeben; und sie sind im Laufe eines langen Evolutionsprozesses zu einem bestimmten Zeitpunkt ‘emergiert’ (aufgetaucht). Folglich sind Physik und Chemie zwar nötig, um die Funktionen des Zentralnervensystems zu erklären, aber sie reichen nicht aus. Auch reicht die allgemeine Biologie nicht aus. Wir müssen vielmehr die spezifischen emergenten Eigenschaften des Zentralnervensystems und die zugrundeliegenden Gesetze erkennen.»

Auch John Searle ist ausdrücklich Monist. Er meint sogar, das logische Rätsel gelöst zu haben, wie Psychisches körperliche Vorgänge bewirken kann: indem es eben nicht, als etwas im Wesen Anderes, in den Körper hineinwirkt, sondern indem es selber etwas Körperliches ist, nur auf einer anderen, höheren Beschreibungsebene betrachtet — so wie Nässe oder Trockenheit keine Eigenschaften sind, die einzelnen Wassermolekülen zukommen, sondern erst zutage treten, wo viele Wassermoleküle zusammenkommen; die Struktur dieses Molekülhaufens «verursache» die Nässe oder Flüssigkeit des Wassers nicht; diese seien vielmehr Eigenschaften dieser Molekülstruktur selber. Ob das Wasser irgendwohin fließt oder Wirbel bildet, läßt sich nicht auf der Ebene seiner Moleküle erklären und von der molekularen Ebene aus vorhersagen. Dennoch beruht es natürlich auf den Eigenschaften seiner Moleküle, wie diese auf ihren atomaren Eigenschaften beruhen. Die Makro-Eigenschaft Nässe geht auf die Mikro-Eigenschaft Molekülstruktur zurück. Einige Phänomene sind nur auf der Makro-Ebene sinnvoll zu beschreiben, andere auf den verschiedenen Mikro-Ebenen. Auf vergleichbare Weise, meint Searle, ergebe sich auch die Makro-Eigenschaft «Psyche» aus den hirnphysiologischen Mikro-Eigenschaften der Gehirnmaschinerie.

Ein schöneres Bild als das des Wassers ist vielleicht das der Wolke. Schon der Oxforder Psychologe Ullian T. Place hat es gebraucht, der 1956 Bewußtsein als einen Gehirnprozeß erklärte und damit jenen Paradigmenwechsel einleitete, der die Mehrzahl der Wissenschaftler inzwischen zu Monisten gemacht hat. Wenn man an die Wolke näher herangeht, verliert sie alles Wolkenhafte und wird zu einem Nebel. Faßt man diesen näher ins Auge, so erkennt man, daß er aus Wassertröpfchen besteht, und die wiederum bestehen aus Wassermolekülen. Einem H_2O -Molekül aber läßt sich nicht entnehmen, wie sich Wassertropfen verhalten, und einem Wassertropfen nicht, was eine Wolke ist. Wie sich eine Wolke türmt und wie sie abschmilzt und wandert, ist nur auf ihrer Makro-Ebene zu erkennen; die reduktionistische Sicht ist nicht direkt falsch, übersieht aber das Wesentliche.

Ein häufiger Einwand gegen die Möglichkeit, irgendwelche Maschinen mit Psyche auszustatten, lautet: Psychisches sei an das biologische Gehirn gebunden und könne nicht entstehen ohne ein biologisches Schicksal, ohne eine individuelle Lebensgeschichte. So sieht es zum Beispiel der

Computerwissenschaftler und -kritiker Joseph Weizenbaum: «Jeder [wurde] von einer Mutter geboren, [hatte] primitive biologische Bedürfnisse ..., [besitzt] einen menschlichen Körper, der beim Verstehen und Wissen zwangsläufig eine Rolle spielt ... Kein Organismus, der keinen menschlichen Körper besitzt, kann diese Dinge in der gleichen Weise wissen wie ein Mensch.» Richtig daran ist, daß so etwas wie Geist oder Verstand bisher ausschließlich an biologischen Gehirnen beobachtet wurde; und daß diese allesamt in Körpern untergebracht sind, die eine Lebensgeschichte haben. Daß es vielleicht auch anders ginge, ist damit aber keineswegs ausgeschlossen. Träumereien wie die des AI-Forschers Hans Moravec, daß der Computer dem Menschen eines Tages Unsterblichkeit bescheren könnte, weil nämlich die individuelle Psyche — das Programm, das von ihr verkörpert wird — auch nach dem Tode des Körpers auf irgendeiner Maschine weiterlaufen könnte, sind heute und gewiß noch auf sehr lange Sicht nichts anderes Science Fiction, und zwar reichlich gewagte. Aber daß sie eine prinzipielle Unmöglichkeit seien: auch das ist bisher nur eine weitere Behauptung.

Ist jemand wie Moravec ein schwarzer Schimmel, ein materialistischer Dualist? In der Tat ist die Kluft zwischen Dualismus und Monismus gar nicht so unüberbrückbar weit, wie die Radikalsten unter ihnen, Physikalisten oder Substanzdualisten, sie erscheinen lassen. Wenn Geistiges zwar aus Materiellem hervorgeht und an keiner Stelle in Widerspruch zu ihm tritt, aber sich aus Materiellem doch nicht restlos erklären läßt, sondern nur auf seiner eigenen Ebene, dann ist er in gewisser Weise doch eine Sache für sich.

Ob der Computer nun eines Tages «Geist» nachmachen kann oder nicht — im Augenblick taugt er auf jeden Fall wie nichts anderes dazu, sich die fast unausdenkbare Beziehung zwischen Körper und Geist andeutungsweise klarzumachen.

Was ist ein Computerprogramm? Es braucht einen materiellen Träger, irgendeinen: einen Streifen gelochten Papiers, eine Scheibe magnetisierter Metallmoleküle. Aber man kann nicht sagen, ein Programm sei nichts als Löcher. Ein Programm ist «Information», eine materielle Repräsentation materieller Sachverhalte (Gegenstände, Vorgänge, Beziehungen). Das Entscheidende daran ist etwas Immaterielles, nämlich eine Ordnung, und wer es nicht auch auf dieser Ebene erfaßt, verfehlt das Eigentliche. Und eben dieses «Immaterielle» steht in Wechselwirkung mit dem Materiellen, «befiehlt» ihm, nimmt Botschaften aus ihm entgegen. An irgendeiner Stelle kommt mein laufendes Computerprogramm, das eine Sequenz elektrischer Signale ist, aber nicht «nur», darauf, daß es die Quadratwurzel aus 26 ziehen oder bestimmen muß, ob eine Zeichenfolge ein Satzprädikat ist. Es «weiß» selber nichts davon, tut es aber, ratiomorph. Auf der Ebene von Siliziummolekülen oder Stromstößen läßt sich nur unzureichend beschreiben und gar nicht erklären, was es da tut, obwohl es dem Charakter von Siliziummolekülen und Stromstößen nirgends widerspricht oder zuwiderhandelt, sondern ganz und gar darauf aufbaut. Es scheint die Mole-

küle herumzukommandieren. Und nun denke man sich noch Spontaneität in den Kasten, eigene Aktivität um eigener Ziele willen, kurz: Leben!

Trägt das Bild ein Stück Wegs, so kann man sagen: Geistige Phänomene beruhen auf Materiellem und verschwinden, wenn die Materie zerfällt, auf der sie beruhen. Aber sie sind nicht *nur* Materie, sondern eine bestimmte, eine dynamische Ordnung, die sich nicht ungestraft reduzieren läßt.

3

Intelligent verhalten werden sich die Computer der Zukunft: ratiomorphe Maschinen. Aber selbst der Computer, die die des Menschen weit hinter sich ließe, wäre, so meine ich, genauso geistlos wie ein Bügeleisen. Er wäre es, weil ihm noch immer eine bestimmte *innere* Fähigkeit abginge: das Bewußtsein.

Bewußtsein – «die offensichtlichste und zugleich geheimnisvollste Tatsache unseres Lebens» (Daniel Dennett) — scheint uns das Selbstverständlichste der Welt, so selbstverständlich, daß wir kaum imstande sind, es aus irgendeinem Leben wegzudenken, dem der eigenen Gattung und dem der eigenen Person schon gar nicht. Dennoch ist jedem geläufig, daß es da sein, aber auch nicht da sein kann. Im Schlaf ist es gedämpft. In der Narkose ist es mutmaßlich völlig ausgeschaltet — wir «merken» nichts mehr, obwohl der Körper und auch seine Leitstelle Gehirn im übrigen weiter funktionieren. Bestimmte Chemikalien trüben und verzerren es. Jeden Morgen geht die Sonne auf, und wenn wir bei Bewußtsein sind, wird es auch im Innern hell. Wir reagieren nicht nur auf irgendwelche elektromagnetischen Schwingungen, wir erleben das Licht. Eigentlich dürfte man nicht sagen: Draußen wird es hell. Draußen nimmt die elektromagnetische Strahlung mit einer Wellenlänge von 400 bis 700 Nanometer zu, die selber gar nicht «aussieht»; erst unser Gehirn ist es, die sie in das subjektive Erlebnis «Licht» übersetzt.

Während das Gehirn Millionen von Operationen gleichzeitig ausführt, wirkt das innere Erleben völlig einheitlich; wir verfolgen die Außenwelt nicht über viele verschiedene Monitore, alles geht in einen einzigen multimodalen Monitor ein. Das innere Erleben weckt auch durchaus das Gefühl, man selber wäre eine Einheit, ein «Ich», daß sich sozusagen das Theater der Welt ansieht, welches ihm die Sinne ins Innere des Kopfes übertragen. Diese integrierende Funktion mag sogar den Hauptvorteil des Bewußtsein ausmachen, den Zweck, um dessentwillen die Evolution es geschaffen hat: Es erzeugt ein einheitliches inneres Modell der Außenwelt, das von einem scheinbar einheitlichen «Ich» erlebt wird. Die tausenderlei visuellen Signale, die ein auf mich zufliegender Ball in meinem Kopf auslöst, werden zusammengefaßt zum Anblick der Flugbahn; und «ich» muß dann nicht Hunderten von Muskeln auf die Millisekunde genau einzelne Bewegungsbefehle zukommen lassen; es genügt, wenn ich mir in meinem

Bewußtsein vorstelle, wie ich mich relativ zu dem Ball bewegen muß, damit ihn mein Schläger trifft; die Umsetzung der Vorstellung in die Befehlssequenz erledigt eine unterbewußte Funktionsschicht des Gehirns.

Man muß sich nur davor hüten, in die alte Homunculus-Falle zu gehen: vor dem Schluß, im Schädel haue eine kleine Person, die sich die Gehirnaktivität betrachtet wie ein Kinobesucher die Leinwand. Denn sofort erhöbe sich die Frage, wie die das macht. Weil auch in der eine noch kleinere Person haue ... und so weiter ad infinitum. Solche Homunculi erklären nichts; sie lassen das, was erklärt werden soll, so unerklärt, wie es war. Im Kopf gibt es keinen Homunculus. Bewußtsein kann nichts jenseits der Gehirnaktivität sein. Es muß eine besondere Dimension dieser Gehirnaktivität selbst sein.

Starrt man einen Begriff wie «Bewußtsein» lange genug an, so zerfließt auch er. Dann wird einem klar, daß man auf verschiedenere Weise bewußt sein kann, daß es «das» Bewußtsein vielleicht gar nicht gibt, daß es zumindest in Graden kommt. Wer eine Wolke einfach nur sieht, ist ihrer auf eine andere Weise gewahr als einer, der seine Aufmerksamkeit auf sie richtet; und wer über sie spricht oder gar etwas über Wolken weiß, in wieder einem anderen Grad. Und sicher die höchste Stufe des Bewußtseins ist, was Selbstbewußtsein heißt und (um Verwechslungen mit dem «gesunden Selbstbewußtsein» der Umgangssprache zu vermeiden) vielleicht besser Eigenbewußtsein hieße: das Bewußtsein der eigenen Person, das Bewußtsein: das bin ich, der dies erlebt. Allen diesen Bewußtseinsmodi aber ist das eine, das Entscheidende gemeinsam: das Gewahrsein, das innere Erleben.

Der heutigen Wissenschaft ist das Phänomen Bewußtsein gar nicht geheuer. Auch in der ganzen Debatte über Computer und Geist kommt es kaum jemals vor. Wenn einer es einmal kurz streift, dann geht er eilig darüber hinweg, als handele es sich nur um eine weitere höhere Hirnfunktion unter vielen, an der ganz und gar nichts Besonderes ist.

Auch John Searle hat erst in einem Aufsatz zum zehnten Geburtstag der von ihm ausgelösten Kontroverse diese seltsame Abstinenz bemerkt: «Nähert man sich der Kognitionswissenschaft, der Psychologie oder der Denkphilosophie mit naivem Blick, so fällt einem als erstes auf, wie wenig Beachtung dem Bewußtsein geschenkt wird. In der Kognitionswissenschaft glauben nur wenige, daß die Erforschung des Geistes im wesentlichen oder zu einem Großteil in der Erforschung von Bewusstseinsphänomenen bestehen müßte; Bewußtsein ist nur ein 'Problem', eine Schwierigkeit, mit der funktionalistische oder computationalistische Theorien irgendwie fertig zu werden haben. Wie aber sind wir in diese Bredouille geraten? Wie können ausgerechnet die Disziplinen, sie sich offiziell der Erforschung des Geistes widmen, sein wichtigstes Merkmal übersehen? Dafür gibt es komplizierte historische Gründe, aber der Hauptgrund ist der, daß wir seit Descartes meist gemeint haben, daß für eine seriöse Wissenschaft oder Denkphilosophie Bewußtsein keinen geeigneten Gegen-

stand bilde. Wenn man noch vor wenigen Jahren in einer Diskussion der Kognitionswissenschaften die Rede auf das Bewußtsein brachte, wurde das allgemein als ein Zeichen schlechten Geschmacks gewertet, und die höheren Semester, immer auf die gesellschaftlichen Mores ihrer Disziplinen ausgerichtet, verdrehten die Augen zur Decke und nahmen eine Miene milden Angewidertseins an.» In einem später konzipierten Vortrag wurde er noch deutlicher: «Die wichtigste Eigenschaft des Gehirns ist es, Bewußtsein hervorzubringen und aufrechtzuerhalten. Es ist der Skandal der Kognitionswissenschaft, gerade sie links liegen gelassen zu haben.»

Einerseits liegt diese Aversion sicher daran, daß die Wissenschaften, auch die *behavioral and brain sciences*, sich sehr viel lieber an objektive Tatsachen halten als an subjektive Erlebnisse; das eine liefert «harte» Daten, das andere solche, für die «weich» noch kein Wort ist. Und daß auch subjektives Erleben eine objektive Tatsache darstellt, diese Einsicht ist darum so überaus unbequem, weil bis heute niemand auch nur die allergeringste Ahnung hat, wie es zustande kommt: wie das Gehirn es fertigbringt, einem kleinen Teil seiner Operationen diese innere Erlebnisqualität zu verleihen, deren gemeinsamen Nenner wir Bewußtsein nennen. Wir wissen mehr über das entfernteste Schwarze Loch als über die elementarste Tatsache unseres Lebens.

Man weiß: Bewußtsein ist an ein entwickeltes Zentralnervensystem gebunden; wahrscheinlich ist es also nicht auf den Menschen beschränkt, möglicherweise beginnt es schon beim Plattwurm heraufzudämmern. Ob auch eine ganz andere Maschinerie etwas Vergleichbares hervorbringen kann, wissen wir schlechterdings nicht.

Man weiß: Bewußtseinsfähig sind nur wenige Körperfunktionen und auch die Hirnfunktionen, die doch Bewußtsein erzeugen, nur zu einem kleinen Teil. Keinem würde noch soviel introspektive Anstrengung dazu verhelfen, sich bewußt zu werden, wie sich die eigenen Pupillen bei einem erfreulichen Anblick weiten, wie das Gehirn Sprachlaute seziert, wie es das Artikulationsprogramm erzeugt, das es an die etwa hundert Muskeln des Stimmtrakts schickt, wenn es beschließt, einen Satz auszusprechen. Es ist einem nicht einmal bewußt, wie man einen bewußten Beschluß faßt; denn das Bewußtsein, daß man etwas beschließt (etwa die Hand zu heben), stellt sich erst etwa eine Drittelsekunde *nach* dem gefaßten Beschluß ein.

Man weiß: Menschliches Bewußtsein setzt die Tätigkeit der Großhirnrinde voraus. Es gibt einseitige Läsionen der Sehrinde, die eine Hälfte des Gesichtsfeldes auslöschen; dort sehen die Kranken nichts mehr — trotzdem reagieren sie weiter auf Dinge, die ihnen auf dieser Seite gezeigt werden. Die Reize also verarbeitet ihr Gehirn weiterhin — nur bewußt wird es ihm nicht.

Man weiß: Wenn der Hirnstamm die Hirnrinde nicht mehr elektrisch aktiviert, erlischt das Bewußtsein. Das Koma ist jener Zustand, bei dem die Retikularformation des Hirnstamms ihre bewußtseinsweckenden Aktivierungen nicht mehr in die Hirnrinde schickt; alle Körperfunktionen,

auch die nicht von Bewußtheit abhängigen Hirnfunktionen gehen weiter, nur das Bewußtsein fehlt.

Man kann trefflich darüber spekulieren, warum die Natur das Phänomen Bewußtsein hervorgebracht hat, welche Vorteile es den Lebewesen bringt, einen inneren obersten Monitor zu besitzen, keine bewußtlosen Reflexautomaten zu sein, die doch auch funktionieren, wenn vielleicht auch nicht so gut. Nur wie das Gehirn den Trick zuwege bringt, das weiß man nicht, und solange dieses Geheimnis nicht gelüftet ist, kann man nicht einmal darüber nachzudenken beginnen, ob und wie auch einem Digitalcomputer Bewußtsein mitzuteilen wäre.

Einer der wenigen, denen aufgefallen ist, daß das Phänomen Bewußtsein ein ernstes Problem für die AI darstellt, ist der Philosoph Daniel Dennett. Aber wie er der AI das Problem vom Hals schafft, ist wenig befriedigend. Er sieht ab von der inneren Erlebnisqualität und definiert das Bewußtsein wiederum als ein äußeres Verhalten: als die Fähigkeit, über seine inneren Zustände zu sprechen. Aber wenn ich — über die so nett sprachverständige Benutzeroberfläche DOS-MAN — meinen Computer frage: «Wieviel Arbeitsspeicher hast du», er darauf in sein Inneres schaut und mir in korrektem Deutsch mitteilt: «29 Dateien belegen 1,2 Megabyte. Frei sind noch 11,6» – dann ist das natürlich kein Beweis für Bewußtsein. (Wie DOS-MAN das macht? Sehr einfach: für all die reichlich abstrakten und trockenen Systembefehle, die sich die Firma Microsoft einst ausgedacht hat, erlaubt DOS-MAN die Definition von Synonymen. Will man sich zum Beispiel ein Verzeichnis aller Dateien auf einem Laufwerk ansehen, so muß man, unter der Standard-Benutzeroberfläche von MS-DOS, den Befehl DIR eingeben. DOS-MAN erlaubt es, diesen wahlweise durch den Befehl ZEIGE DATEIEN oder jede andere Zeichenfolge, die einem zusagt, zu ersetzen. Hinfort zeigt der Computer seinen Inhalt auch dann, wenn man, statt DIR, ZEIGE DATEIEN tippt; und wenn man will, auch so etwas wie ZEIGE MIR BITTE SOFORT ALLE HEUTIGEN DATEIEN MIT DER ENDUNG TXT. Die Antworten, die er einem gibt, sind alle vorformuliert; der Computer trägt nur gegebenenfalls die aktuellen Werte ein. DOS-MAN «versteht» gar nichts und hat natürlich keinerlei Bewußtsein; im Kern ist es nur ein — sinnreich ausgedachter — Mechanismus zum Vergleichen von Zeichenketten. So einfach ist es, Dennetts Kriterium für Bewußtsein zu genügen.)

Fast als einziger sieht John Haugeland, Philosophieprofessor an der Universität Pittsburgh, daß da eine große beunruhigende Unbekannte ihr Wesen treibt: «Bewusstsein — das ist ein Thema, welches in der Fachliteratur der Kognitionswissenschaft auffällig abwesend ist. Es liegt nahe, daß solch ein dröhnendes Schweigen ein häßliches kleines Geheimnis birgt ... Könnte das Bewußtsein eine theoretische Zeitbombe sein, die im Schoß der AI tickt? Wer weiß?»

Ich meine also: Searle im Chinesischen Zimmer hat völlig recht und völlig unrecht. Recht hat er, wenn er sagt, daß der Computer nur dann so

etwas wie menschlichen Geist haben könnte, wenn er genau jene Eigenschaften besäße, die dessen wesentliche Attribute verursachen. Unrecht hat er, wenn er in der Intensionalität sein wesentlichstes — und zur Zeit unnachahmbares – Attribut sieht. Nicht die Intensionalität seines Geistorgans trennt den Menschen vom Computer, sondern seine Bewußtseinsfähigkeit. Wenn das nicht allgemein klar ist, dann nur, weil manch einer sich Intensionalität stillschweigend so zurechtdefiniert, daß sie Bewußtheit einschließt, weil er einen intensionalen Akt gar nicht anders als bewußt denken kann. Das kann er aber sehr wohl sein: Die meisten Funktionen des Zentralnervensystems – etwa der automatische Abzug von Blau vom Morgen- und Rot vom Abendlicht oder die automatische gegenläufige Augenbewegung zum Ausgleich von Kopfbewegungen — sind intensional, gelangen aber nicht zu Bewußtsein. Inzwischen sieht wohl auch Searle, daß das Entscheidende nicht die Intensionalität, sondern das bewußte innere Erleben ist. Wenn wir eine Flasche ohne Skrupel zerschmettern, aber eine Katze nicht, dann einzig weil wir (letztlich unbeweisbar, aber sehr wahrscheinlich zu Recht) vermuten, daß diese das als Qual empfindet, also ein Bewußtsein davon hätte. Solange wir nicht wüßten, daß ihm Gewahrheit, Empfindungsfähigkeit, Bewusstsein — ein inneres Erleben — eingebaut sind, zögen wir auch dem intelligentesten Computer ohne das mindeste Zögern herzlos den Stecker aus der Dose.

Der Hinweis auf das völlig ungeklärte Phänomen «Bewußtsein» ist nicht als die Neuauflage irgendeines Dualismus zu verstehen. Auch das Bewußtsein wird nichts Außernatürliches sein, sondern eine emergente Funktion des materiellen Gehirns, die wie alle anderen Hirnfunktionen objektiver Untersuchung zugänglich ist. Nur daß wir eben leider partout nicht wissen, welche Eigenschaften der organisierten Materie die Natur ausgenutzt hat, dieses Wunder hervorzubringen.

Wenn die Kognitionswissenschaft die riesige Kluft zwischen der höchsten Ebene der geistigen Funktionen wie «Denken» oder «Sprechen» oder «Rechnen» und der untersten Ebene der Hirnphysiologie mit ihren «feuernden» Neuronen zu überbrücken versucht, bedient sie sich gemeinhin der Idee der «rekursiven Dekomposition»: Sie zerlegt die geistige Funktion immer und immer wieder auf die gleiche Weise, bis sie auf ein paar primitive Operationen zurückgeführt ist — genau so, wie es der Computer täte. Ganz oben multipliziert jemand intelligent 3 und 3; auf der Ebene darunter wartet das Gehirn mit einem dümmen Mechanismus auf, der zwar nicht multiplizieren, aber addieren kann und die Aufgabe 3 mal 3 in 3 plus 3 plus 3 verwandelt. Darunter ist ein Mechanismus, der auch keine 3 mehr kennt, sondern nur noch zwei Zahlen. Und ganz unten befinden sich ganz dumme einzelne Neuronen, die nur noch «An» und «Aus» sagen können, wie der Computer im Innersten nur 0 und 1 schreibt. Für die komputatorischen geistigen Fähigkeiten scheint das auch ganz plausibel; sie lassen sich auf diese Weise sezieren, der Computer macht es vor. Die Bewußtheit aber läßt sich so überhaupt nicht dekomponieren. Sie scheint eine nichtglobale Fähigkeit, die sich nicht logisch auf immer schlichtere

Fähigkeiten zurückführen läßt. Vielleicht nur, weil wir ihr Wesen noch so gar nicht durchschauen; aber vielleicht rührt diese Unmöglichkeit auch von ihrer Natur her.

Ein anderes populäres Gedankenexperiment auf diesem Gebiet ist das folgende. Man stelle sich vor, man könnte winzige Prozessoren bauen, die genau das tun, was die einzelnen Neuronen des Gehirns tun: elektrische Potentiale aufzunehmen, zu addieren, zu subtrahieren und selber welche abzugeben, wenn eine bestimmte Summe erreicht ist. Es scheint nicht grundsätzlich unmöglich. Und nun beginne man, die Neuronen des Gehirns eins nach dem andern durch die entsprechenden Prozessoren zu ersetzen, bis man am Ende einen dichtverdrahteten Siliziumfilz hat, der Struktur und Funktion des Gehirns vollkommen nachbaut. Wäre er eine Gehirnprothese? Würde er denken wie das Gehirn? Hätte er Bewußtsein?

Ich glaube, die Antwort kann heute nur lauten: Nur dann, wenn bei der Prozedur nicht unabsichtlich genau jene Eigenschaft des Gehirns abhanden gekommen ist, die Bewußtsein hervorbringt. Um sie zu erhalten, müßte man aber erst einmal wissen, worin sie besteht, und da man das nicht im mindesten weiß, kann man auch nicht das mindeste tun, einem Computer zu bewußtem Geist zu verhelfen.

Literatur

- Bertalanffy, Ludwig von:** «The Mind-Body Problem – A New View». In: *Psychosomatic Medicine*, 26, 1964, S.29-45
- Braitenberg, Valentino:** «Brains and Computers». In: Kenneth Klivington: *The Science of Mind*. Cambridge, MA: MIT Press, 1989, S.229-230
- Bunge, Mario:** *The Mind-Body Problem*. Pergamon Press, Oxford 1980. Deutsch: *Das Leib-Seele-Problem – Ein psychobiologischer Versuch*. Tübingen: Mohr, 1984
- Changeux, Jean-Pierre:** *L'homme neuronal*. Fayard, Paris 1983. Deutsch: *Der neuronale Mensch*. Reinbek: Rowohlt, 1984
- Churchland, Patricia Smith:** *Neurophilosophy – Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press, 1986
- Churchland, Paul M. / Patricia Smith Churchland:** «Künstliche Intelligenz II – Ist eine denkende Maschine möglich?» In: *Spektrum der Wissenschaft*, 3, März 1990, S.47-54
- Churchland, Paul M.:** *Matter and Consciousness – A Contemporary Introduction to the Philosophy of Mind*. Cambridge, MA: MIT Press, 1984

- Dennett, Daniel C.:** *Brainstorms – Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press, 1981
- Dreyfus, Hubert L.:** *What Computers Can't Do – The Limits of Artificial Intelligence*. New York: Harper & Row, ²1979. Deutsch: *Was Computer nicht können – Die Grenzen künstlicher Intelligenz*. Frankfurt/Main: Athenäum, 1989
- Eccles, Sir John:** *Evolution of the Brain – Creation of the Self*. London: Routledge, 1989. Deutsch: *Die Evolution des Gehirns – die Erschaffung des Selbst*. München: Piper, 1990
- Feigl, Herbert:** «The Mind-Body Problem: Not a Pseudo-Problem». In: Sidney Hook (Hg.): *Dimensions of Mind*. New York: New York UP, 1960
- Furst, Charles:** *Origins of the Mind – Mind-Brain Connections*. Englewood Cliffs, NJ: Prentice-Hall, 1979
- Gierer, Alfred:** *Die Physik, das Leben und die Seele*. München: Piper, 1985
- Gray, Jeffrey:** «The Mind-Brain Identity Theory as a Scientific Hypothesis: A Second Look». In: Colin Blakemoore/Susan Greenfield (Hg.): *Mind-waves*. Oxford: Blackwell, 1987
- Harnad, Stevan:** «Minds, Machines and Searle». In: *Journal of Experimental and Theoretical Artificial Intelligence*, 1 (1), 1989, S.5-25
- Haugeland, John:** *Artificial Intelligence – The Very Idea*. Cambridge, MA: MIT Press, 1985
- Hofstadter, Douglas R. / Daniel C. Dennett (Hrsg.):** *The Mind's I – Fantasies and Reflections on Self and Soul*. New York: Basic Books, 1981
- Minsky, Marvin:** *The Society of Mind*. New York: Simon & Schuster, 1985. Deutsch: *Mentopolis*. Stuttgart: Klett-Cotta, 1990
- Moravec, Hans:** «Künstliche Intelligenz – Der Mensch als Maschine». In: *DIE ZEIT/ZEITmagazin*, 12, 16. April 1990, S.28-34
- Place, Ullian T.:** «Is consciousness a brain process?» In: *British Journal of Psychology*, 47, 1956, S.44-51
- Pribram, Karl:** «Consciousness and neurophysiology». In: *Federation Proceedings*, 37 (9), 1978, S.2271-2274
- Schank, Roger C. / R. P. Abelson:** *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum, 1977
- Searle, John R.:** «Minds, brains, and programs». In: *Behavioral and Brain Sciences*, 3, 1980, S.417-457 [mit Diskussionsbeiträgen]
- Searle, John:** *Minds, Brains and Science*. Cambridge, MA: Harvard UP, 1984

- Searle, John R.:** «Künstliche Intelligenz I – Ist der menschliche Geist ein Computerprogramm?» In: *Spektrum der Wissenschaft*, 3, März 1990, S.40-47
- Searle, John R.:** «Consciousness, Explanatory Inversion and Cognitive Science». In: *Behavioral and Brain Sciences* [1990; in Vorbereitung]
- Sperry, Roger W.:** «Mental Phenomena as Causal Determinants in Brain Function». In: Gordon G. Globus/Grover Maxwell / Irwin Savodnik (Hg.): *Consciousness and the Brain – A Scientific and Philosophical Inquiry*. New York: Plenum Press, 1976
- Turing, Alan M.:** «Computing Machinery and Intelligence». In: A. R. Anderson (Hg.): *Minds and Machines*. Englewood Cliffs, NJ: Prentice-Hall, 1964, S.4-30
- Weizenbaum, Joseph:** «Künstliche Intelligenz – Absurde Pläne». In: *DIE ZEIT/ZEITmagazin*, 12, 16. April 1990, S.38-41